LORELEI

Low Resource Languages for Emergent Incidents

Proposers' Day

13 November 2014

Boyan Onyshkevych

DARPA

Approved for Public Release, Distribution Unlimited

LORELEI Proposers' Day Agenda

Start	End	Session
08:30	10:00	Registration
10:00	10:45	Opening Briefing (Attendees submit questions)
10:45	11:00	CMO Briefing (Attendees submit questions)
11:00	11:10	Security Briefing (Attendees submit questions)
11:10	12:30	Attendees may speak for 2 minutes for teaming purposes. No slides or handouts. DARPA representatives will not be present.
12:30	01:15	Lunch – on your own
01:15	02:45	Question Answering Session
02:45	7:00	5-Minute one-on-one meetings with LORELEI program manager

LORELEI Proposers' Day Logistics

- BAA Location and Dates
  - Posted on FedBizOpps website (http://www.fedbizopps.gov) and Grants.gov website (http://www.grants.gov)
  - Posting Date: October 30, 2014
  - BAA Closing (Proposal Due Date): January 9, 2015, 12:00 noon (ET)
- Procedure for Questions/Answers Today
  - Questions can be submitted until 11:30 to LORELEI@darpa.mil or on 3x5 cards
  - Questions will be answered during Q&A session in the afternoon
- Websites
  - Proposers' Day website
  - LORELEI program website
    - Proposers' Day Presentations
    - Frequently Asked Questions (FAQ) will be updated with Q/A from LORELEI@darpa.mil
- Information precedence
  - If anything said or addressed during this presentation or in the FAQ conflicts with the published solicitation, the BAA takes precedence. The Government may issue amendments to the BAA to effect any changes deemed necessary in response to the FAQ. Such amendments would be posted to FBO and Grants.gov prior to the solicitation closing date and would supersede previous versions of the solicitation.

LORELEI Program Manager Presentation Outline

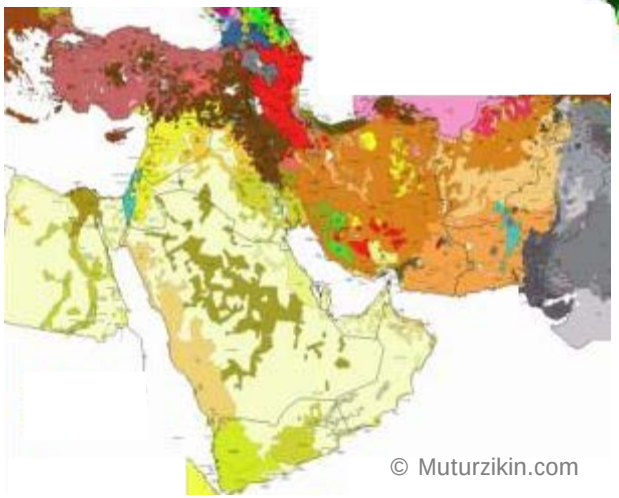
- Program Goal
- CONOPS
- Program Structure
  - Technical Area 1: Algorithm Research and Development Environment
  - Technical Area 2: Run-time Framework Development
  - Technical Area 3: Linguistic Resource Creation
  - Evaluation
- Expected Program Schedule
- Proposal Information
- Reference Documents

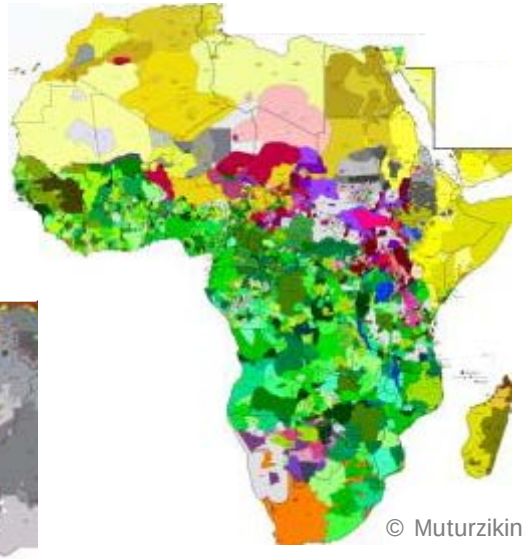
Program Goal: Rapid Response Language Technology

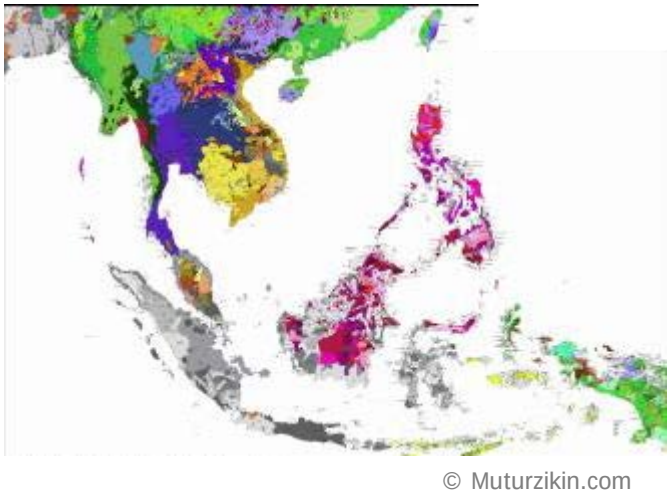
Rapid development of language technology to provide situational awareness based on information from **ANY** language, in support of emergent missions

LORELEI success will mean non-linguist mission planners can achieve results comparable to expert linguists'

Language Variation
(Colors represent language groups)





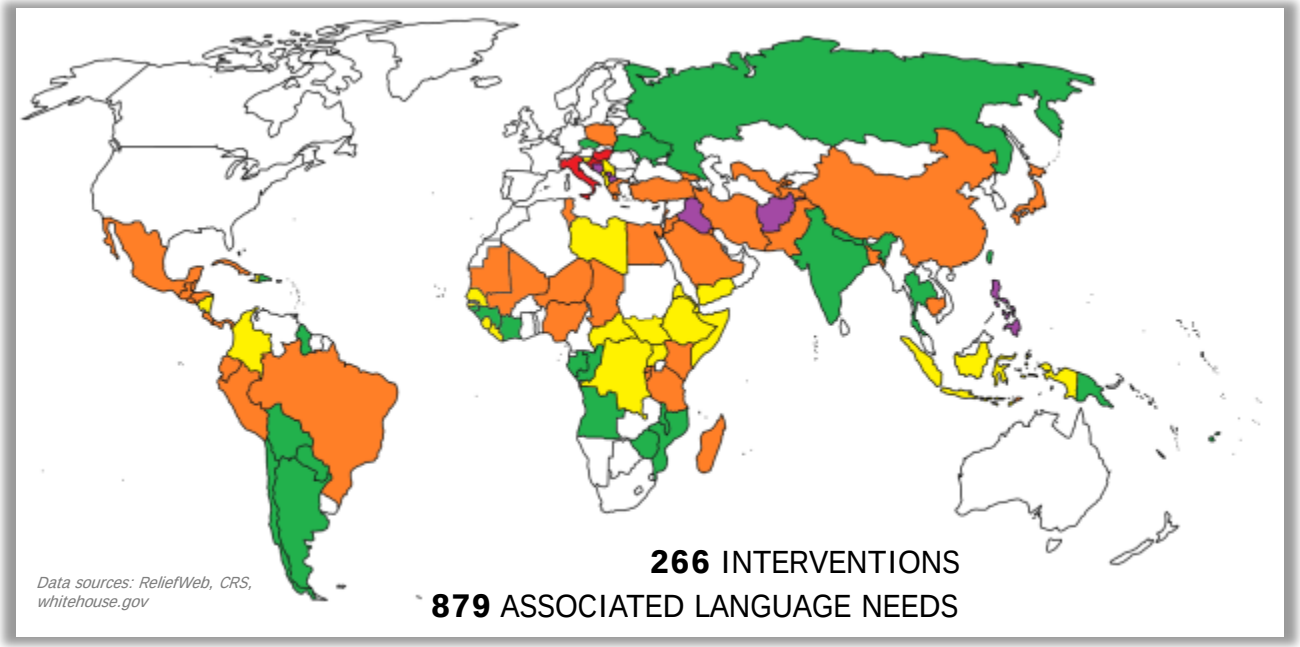


Rapid Response Language Technology -
Initial capability beginning within **24 hours** of emergent need

Program Goal: Why LORELEI?

- 7100+ active languages in the world - hard to predict which languages will be needed next
  - 44 in Boko Haram area (Hausa, Kanuri) – 522 languages in all of Nigeria
  - 19 in Ebola outbreak areas in Liberia, Sierra Leone, and Guinea (Kissi, Kuranko)
  - 20+ Mayan languages spoken by Central American refugee children (Q'anjob'al, K'itche, Ixil)
- Developing language technologies by current methods requires approximately 3 years and 10s of millions of $ per language (mostly to construct translated or transcribed corpora)
  - By current methods, need $70B and 230,000 person-years of effort for all languages

Language Requirements Example: Foreign Disaster Relief Language Needs, 1990-2014



Only 22 of these languages with current machine translation and/or information extraction capabilities

Program Goal: What do we mean by “Low Resource”?

- For the purpose of this program, “Low Resource” is used to refer to the lack of availability of mature Human Language Technology
- So most languages in the world will be “Low Resource” for the purpose of LORELEI, with the exception of at most several dozen languages
- Exact list of what is or isn’t a “Low Resource” language is not important, since the program will be exercising technologies for processing languages relevant to incidents at a range of resource levels
- “Low Resource” does not refer to the size of the population speaking that language
- LORELEI languages may be from anywhere in the world, especially Africa and Asia, and may be reflective of ongoing crisis incidents

Approved for Public Release, Distribution Unlimited

Hypothetical HADR Application Example

An earthquake-tsunami event results in the destruction of all water villages and many urban areas in a coastal country.

•**1 DAY** - Within one day of the disaster, info identified by LORELEI from low resource language media cue responders to large pockets of entrapped peoples and enables intelligent (rather than ad hoc) triage of locations requiring rescue and evacuation.

•**1 WEEK** - Devastation and lack of food in certain areas causes thousands of survivors to seek food and shelter in the mountains to the east. One week into the disaster, LORELEI cues responders to what are becoming de facto evacuation centers and alerts responders to the requirements at these locations. Info identified by LORELEI enables prioritization of needs and road/path networks to supply those needs.

•**1 MONTH** – Tribal communities, resentful of the sudden incursion of urban peoples, become increasingly frustrated and begin organizing to obtain arms with which to eject the urban refugees. LORELEI enables responders to become of aware of the changes in people and organizations and to prioritize the de facto evacuation centers in most need of mitigating actions.

Program Goal: Rapid Response

Scenario Types

- Humanitarian Assistance and Disaster Relief (HADR)
- Peacekeeping, Counterterrorism, Law Enforcement Assistance
- Medical Aid for Infectious Disease Outbreaks, Radiological Incidents

Example Output: Hotspot Map
© Ushahidi

Response Timeline

- **Hotspots**: places where human or material resources are needed – 1 Day
- **Logistics**: location and status of transportation routes - 1 Week
- **People and Organizations**: human activities - who, what, when, where – 1 Month

	1 Day	1 Week	1 Month
Place Names	●	●	●
Topic Spotting in Text	◑	●	●
Sentiment/Emotional State	◑	●	●
Partial Translation	◑	◑	●
Events	◑	◑	◑
Relationships	○	◑	●
Person Information	○	◑	●
Topic Spotting in Speech	○	◑	◑
Full Translation	○	◑	◑
Speech to Text	○	○	◑

Create tools for translating incident language and for identifying elements of information in incident language and English sources

Expose information to analyst (or to analytics) via web services

Approved for Public Release, Distribution Unlimited

CONOPS: Hypothetical Mission Requirements and Answers

	Mission Requirement	Questions Answered by LORELEI (Ebola outbreak scenario)
Limited Characterization of the Environment	Discover what locations are involved	• Where are there confirmed and suspected cases? • Where are reactions to the outbreak occurring?
	Rate the urgency at each location	• How many cases are there at each location? • What level of violence is involved in reactions?
Extended Characterization of the Environment	Ascertain the type of response needed for each location.	• What types of shortages are occurring? • What types of people are infected? • When did the infections occur?
	Determine what personnel and supplies are required	• What are the available local resources and facilities? • Which organizations are already on the ground? • What resources are needed? Medicine, food, water?
	Plan for coordination with local people and organizations	• Who is in charge of relevant organizations? • Who are the political officials or de facto leaders?
	Map out routes for forces to get to each location	• What transportation is available? • What routes are impeded by quarantines, roadblocks, or violence?
Full Characterization of the Environment	Determine what containment measures are required	• Where are sick people and refugees moving? • Where is the outbreak under control/not under control?
	Plan for adjustments in response to changes in circumstances	• What is the reaction of the infected people? The uninfected? • Is the quarantine causing unrest? Violence? • Is there an increased level of crime?

CONOPS: Timeline

- $T_{ø}$: A language need emerges
  - Analysts must be able to produce mission plans, situational awareness documents, or other analytic products for crisis responders. No language processing technology is available for the relevant language.
- $T_{ø}$ + 1 day: initial analyst/mission planner interaction with LORELEI
  - Analysts must be able to use LORELEI to create an initial mission plan in one hour, for example a list identifying specific hotspots most relevant to the crisis or locations requiring the greatest assistance.
- $T_{ø}$ + 1 week : increased analyst/mission planner interaction with LORELEI
  - Analysts must be able to use LORELEI to produce enhanced mission plans within three hours, specifying, for example, viable transportation or supply routes, means of avoiding blocked roads, etc.
- $T_{ø}$ + 1 month : full analyst/mission planner interaction with LORELEI
  - Analysts must be able to use LORELEI-provided information to develop an understanding of the whole situation in the crisis area, representing organizations, involved people, etc. in about three hours of interaction with LORELEI.

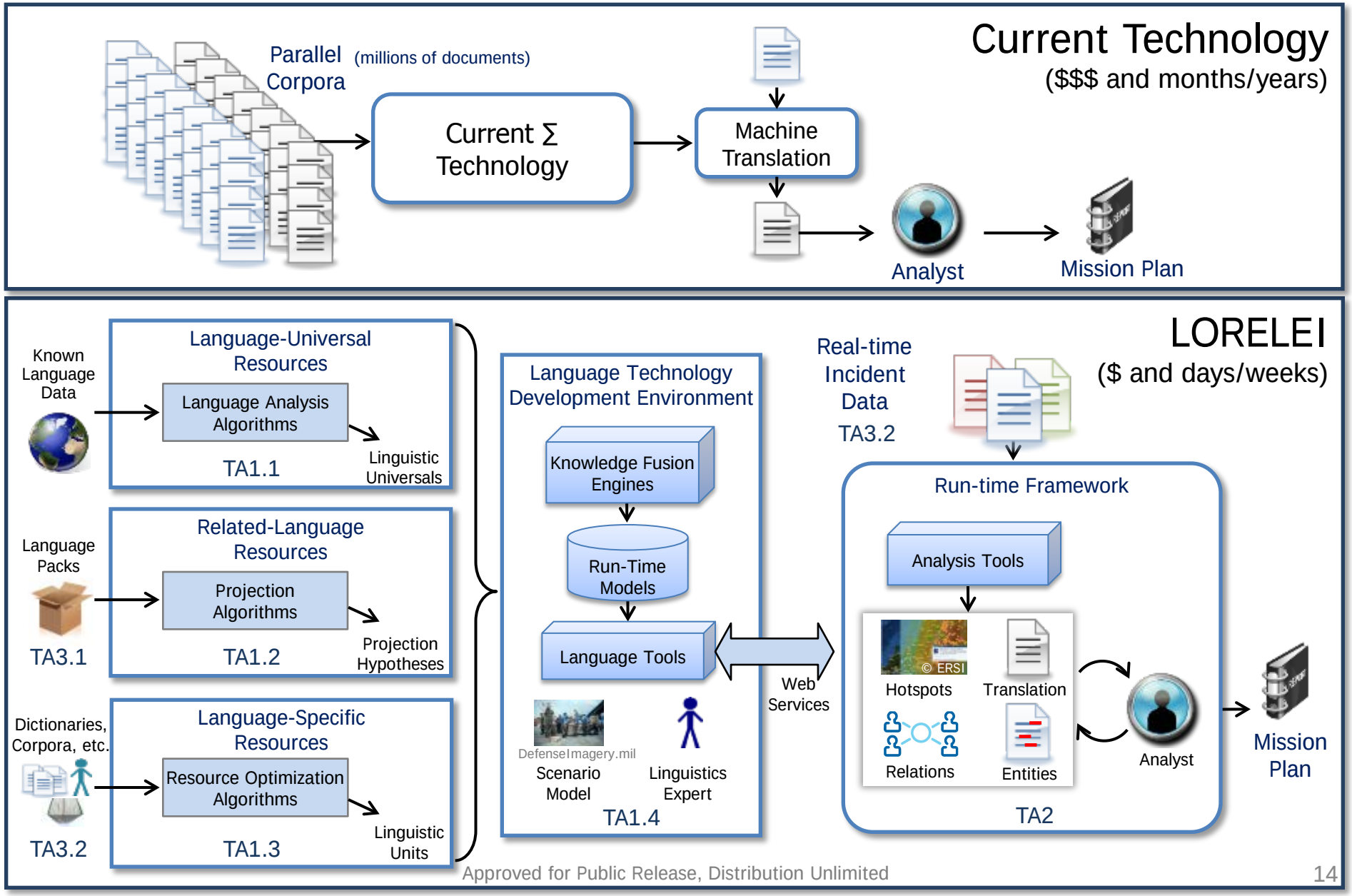
CONOPS: Inputs

- Language acquisition ("training") resources
  - Monolingual corpora
    - Archival incident language data in multiple genres
  - Parallel incident language-English corpora
    - Archival parallel corpora in the incident language and English or other language
  - Parallel dictionary
    - English and the incident language
  - Linguistic resources of variable type, size, and quality
    - Dictionaries, grammar books, gazetteers, etc.
  - LORELEI representative language packs (LRLPs)
    - Resource sets for typologically representative languages
  - Native-speaker linguistic informant
    - Access to a certain number hours with a native-speaker linguistic informant
  - Expert theoretical linguist
    - Access to unlimited hours of one on-site expert theoretical linguist
  - Scenario model in English
    - Scenario model specific to incident type
- Incident-related input data
  - Streaming incident language data (incident-related topics mixed with other topics) including news and informal genres
  - Streaming English newswire (incident-related topics mixed with other topics)

CONOPS: Output

- Web services that process incident-related IL data
  - Required capabilities:
    - Named-entity recognition (for foreign language text)
    - Term/topic discovery and classification (for foreign language text and speech)
    - Machine translation (for foreign language text)
  - Optional additional capabilities, such as:
    - Named-entity coreference identification
    - Entity attribute identification
    - Event extraction
    - Relation extraction
    - Topic modeling
    - Sentiment/emotion detection
    - Etc.
  - While some of these capabilities may be based on partial or full transcription and/or translation, the goal is providing information to support situational awareness, not just English translations of foreign language sources.
- Additionally, LORELEI is expected to produce Web services that provide named-entity recognition, topic spotting, and several other language technology capabilities for English-language newswire.
- Run-time framework that calls the above web services and supports the user in developing the mission plan/situation awareness

Program Structure: LORELEI Concept



Approved for Public Release, Distribution Unlimited

Program Structure: Technical Areas

Technical Area 1 Algorithm Research and Development Environment

- Research and development of algorithms and/or tools in the four focus areas:
  - TA1.1: Leveraging Language-Universal Resources
  - TA1.2: Projecting from Related-Language Resources
  - TA1.3: Exploiting Language-Specific Resources
  - TA1.4: Language Technology Development Environment

Technical Area 2 Run-time Framework Development

- Development of a task-specific run-time CONOP and prototype that accepts various data streams, calls TA1-built language-specific web services, and runs analysis tools on processed data to aggregate, summarize, and organize processed information

Technical Area 3 Linguistic Resource Creation

- Collection, creation, and annotation of linguistic resources including representative language packs and research data to support research, development, and evaluation
  - TA3.1 - Representative language pack creation
  - TA3.2 - Incident language resource creation to support research and evaluation

Approved for Public Release, Distribution Unlimited

TA1.1: Leveraging Linguistic-Universal Resources

- Focus: research and development of novel techniques for discovering and utilizing "universal" properties and (typological or other) regularities of language, or exploiting communication or information theory to constrain search space for learning language models.

- Objective: build on what is known or can be discovered about the characteristic tendencies and regularities of human language, but not limited to "absolute" universals that apply to every known language.

- Results: results must not be solely theoretical in nature; they must provide substantive support to the overall program goal and contribute to run-time incident language tools in the Language Technology Development Environment (LTDE).

Example "universal" linguistic patterns:

- Syntax: In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object. (Greenberg 1966)

- Morphology: If the verb has categories of person-number or if it has categories of gender, it always has tense-mode categories. (Greenberg 1966)

- Phonology: no language has b/d/g without p/t/k (Jakobson 1968)

Approved for Public Release, Distribution Unlimited

TA1.2: Projecting from Related-Language Resources

- Focus: novel techniques for projecting from resources and models for a fixed set of representative languages to another language (i.e., to the incident language). To support TA1.2, LORELEI Representative Language Packs (LRLPs) will be constructed under TA3.1 for 24 languages, spanning a range of language families.

- Objective: development of cross-language projection techniques from the LRLPs to incident languages.

- Results: results must not be solely theoretical in nature; they must provide substantive support to the overall program goal of enabling situational awareness acquisition based on incident language sources.

Example data set similar to an Uzbek LRLP available now

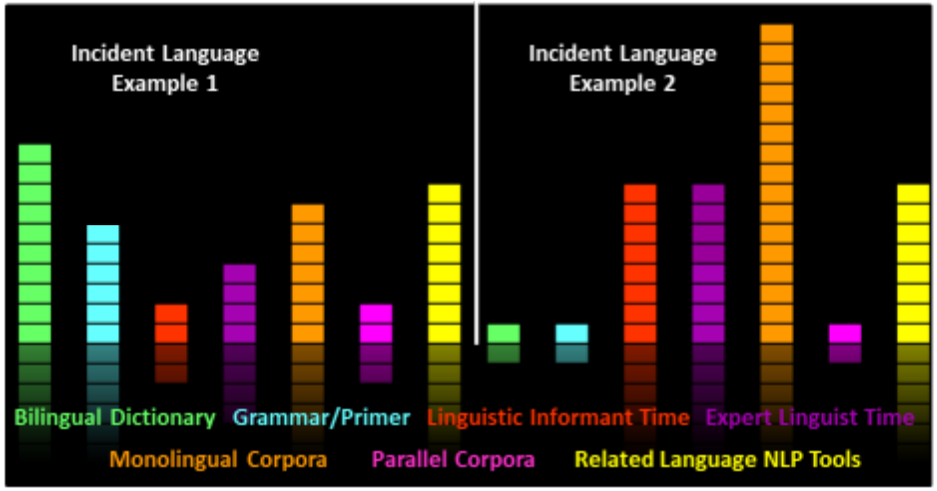
Language	"nine"	"mountain"
Turkish	dokuz	dağ
Azeri	doqquz	dağ
Uzbek	toqqiz	toğ
Uyghur	toqquz	tağ
Tatar	tuğız	taw
Kazakh	toğız	taw
Kyrgyz	toğuz	tō
Altay	toğus	tū
Fu-yü Gyrgys	doğus	daχ
Khakas	toğıs	tağ
Tuvan	tos	dağ
Sakha/Yakut	toğus	tıa
Khalaj	toqquz	tāğ

TA1.3: Exploiting Language-Specific Resources

- Focus: novel techniques for maximizing the utility of existing linguistic resources in the incident language. Possible tasks include:
  - Developing methods for improving use of pre-existing resources
  - Determining the optimal balance of resources that are available on a case-by-case basis
  - Utilizing a native-speaker linguistic informant efficiently and (limited) crowdsourcing

- Objective: determine the answers to questions like:
  - What should the linguistic informant do in limited time – transcribe and/or translate words/sentences, mark topics/names, diagram sentences?
  - How can more information can be derived from each resource and reliance on huge parallel corpora be reduced?
  - Can frequently-spoken short speech segments be discovered and utilized?
  - How to improve unsupervised, distantly-supervised, or very-lightly-supervised methods to the point of being useful for the LORELEI concept of operations?

- Results: results of work in TA1.3 must provide significant improvement on existing resource utilization, creation, or exploitation techniques, not consist of just providing resources.

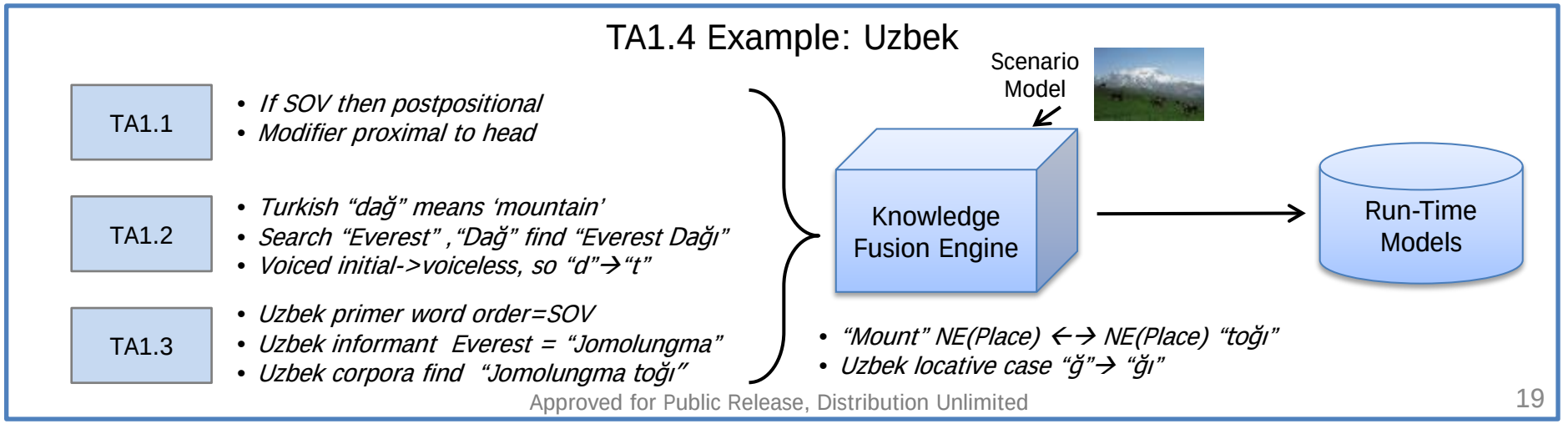


TA1.4: Language Technology Development Environment

- Focus: novel techniques for integrating algorithms, tools, and linguistic information from TAs 1.1-1.3 into a single development environment (the LTDE). The LTDE:
  - Fuses partial incident language information from TA1.1-1.3 into run-time models
  - Exposes language-processing tools with incident-language models at runtime via Web services.
  - Provides (via Web service) existing English language processing tools for newswire, including Named-Entity Recognition, topic spotting, and other functions as needed.

- Objective: Fuse the models, rules, assertions, or extracted information produced by TA 1.1, 1.2, and 1.3 algorithms to create run-time models for the program-specified incident languages, running on TA1-produced language tools.

- Results: All proposals addressing TA1.4 must also address TA1.1, TA1.2, and TA1.3 (i.e., each TA1.4 proposal must address all of TA1).



TA1.4: Performer Information

- DARPA will provide to TA1.4 performer (if desired):
  - English named entity tagger (beginning of program)
  - English topic spotter (beginning of program)

- TA1.4 proposers are encouraged to assemble teams as needed to address the entirety of TA1

TA2: Run-time Framework Development

- Access TA1-built run-time incident language processing tools via Web services
- Run analysis tools on processed data to aggregate, summarize, and organize information on the incident
- Build on existing tools for guided search, visualization, fusion, etc. (e.g., DARPA XDATA, Memex, Insight)
- TA2 performer not responsible for development of any language tools.

Real-time Incident Data

Run-time Framework

Run-Time Models

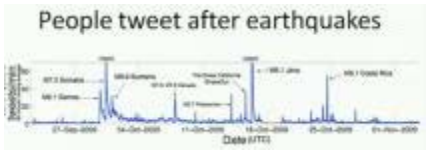
Language Tools

Web Services

Analysis Tools

Hotspot Map

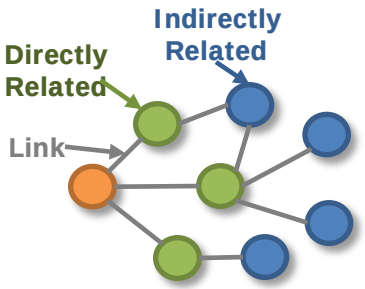
Events



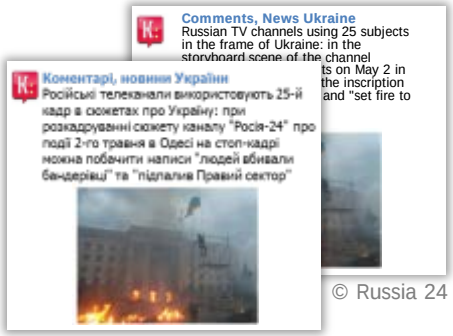
© Paul Earle

OUTPUTS

Social Graph



Translation



© Russia 24

Analyst

Mission Plan

Approved for Public Release, Distribution Unlimited

TA2: Performer Information

TA2 Performer needs to have cleared system engineers

TA2 Performer expected to travel to CONUS and OCONUS facilities of COCOM transition partners/beta sites

TA3: Linguistic Resource Creation

- TA3 will be focused on collection, creation, and annotation of linguistic resources supporting the work of TA1 and TA2. These resources will include the LRLPs and other data enabling research, development, and evaluation.
  - TA3.1 focus: creation of LRLPs in languages designated by the Government. DARPA will provide previously-collected or created materials to partially complete LRLPs for certain languages.
  - TA3.2 focus: collection and/or creation of linguistic data resources to support the research and development conducted under TA 1 and TA2 and to support evaluation.
- There may be different performers for TA3.1 and TA3.2. Multiple partial awards for each are possible.

TA3.1: LRLP Contents

Monolingual Data	**2,100,000**
Newswire	1,000,000
Blogs	400,000
Discussion forums	400,000
Microblogs	200,000
Speech (formal / informal, as available)	100,000
Parallel Data (translated into English)	**900,000**
Newswire	450,000
Blogs	180,000
Discussion forums	180,000
Microblogs	90,000
Parallel Data (translated from English)	**110,000**
REFLEX English core set*	66,000
REFLEX elicitation corpus*	23,000
REFLEX phrase book*	8,300
Additional in-domain English newswire	12,700

Annotated Data	**265,000**
POS tagged gold standard text	10,000
Morph analyzed gold standard text	10,000
Morph-level alignment of parallel text	10,000
NP chunking for text	10,000
Simple NE tagged text	75,000
Full entities w/ coref for text	25,000
Simple semantic annotation of text	25,000
Topic annotated speech	100,000
Annotation guidelines	n/a
Miscellaneous Items	**n/a**
POS, morph tagset	n/a
Specialized word lists (GPEs, etc.)	n/a
Grammatical sketch	n/a
Lexicon	10,000
Encoding converter	n/a
Sentence segmenter	n/a
Word segmenter	n/a
POS tagger	n/a
Morphological analyzer	n/a
Named entity tagger	n/a
Name transliterator	n/a

*Prior REFLEX program involved selection/creation of a common set of data to be translated from English into target languages, elicited from target language speakers, and translated from each target language into English.

TA3.1: LRLP Languages

- Expected languages of LORELEI Related Language Packs (LRLPs):

LORELEI Representative Language Packs			
Akan (Twi)	Guarani*	Russian	Thai*
Amharic*	Hausa*	Somali	Uzbek*
Arabic*	Hindi	Spanish	Vietnamese
Bengali*	Hungarian*	Swahili	Wolof
Burmese*	Indonesian	Tagalog*	Yoruba*
Farsi	Mandarin*	Tamil*	Zulu

*Data will be provided by DARPA to partially complete LRLPs for these 13 languages

- LRLP1 Uzbek basically complete version has already been released.
- LRLP2 Hausa basically complete version will be released at the start of the program
- LRLPs3-13 will be released approximately 12 months after program start
- LRLPs14-24 will be released approximately 24 months after program start

TA3.2: Incident Language Resources

For all Incident Languages (includes both "training" and incident data)

- Monolingual corpora
  - Archival incident language data
  - Streaming incident language data (incident-related topic mixed with other topics)
  - Streaming English data including (incident-related topic mixed with other topics)
- Parallel incident language-English corpora
  - Archival parallel corpora in the incident language and English or other language
- Parallel dictionary
  - Incident Language <-> English
- Linguistic resources of variable size and quality (at least 5 of these 8 items for each IL)

• Parallel dictionary	(IL->non-English)	Monolingual primer book	(IL)
• Monolingual dictionary	(IL)	Monolingual Gazetteer	(IL)
• Monolingual grammar book	(IL, may be hardcopy)	Parallel Gazetteer	(IL-English)
• Parallel grammar book	(IL->English, may be hardcopy)	English Gazetteer	(incident region)

For Evaluation Incident Languages Only (and one sample IL to be released early)

- Annotated Text Corpora (MT metric reference, NE tagged, topic labeled)
  - Formal text
  - Informal text
  - Text communication
- Annotated Speech Corpora (topic labeled)
  - Formal speech
  - Speech communication

TA3.2: Incident Languages

- LORELEI languages may be from anywhere in the world, especially Africa and Asia, and may be reflective of ongoing crisis incidents.
- There will be two languages designated for development and one “surprise” language chosen for evaluation every 12 months.
  - IL1 = Turkish. A basically complete version has already been released.
  - IL2 = Mandarin. A basically complete version will be released at the start of the program.
  - IL3 = first evaluation language. It will not be announced to TA1 and TA2 performers until the evaluation, 12 months after program start.
  - The other development ILs (IL4, IL5, IL7, IL8, IL10, IL11) will be announced incrementally, as the program proceeds, and data will be delivered to TA1 and TA2 performers two ILs at a time, in 12 month increments.
  - The other evaluation ILs (IL6, IL9, IL12) will not be announced to TA1 and TA2 performers until each evaluation.
  - The first evaluation IL will be related to the language of at least one LRLP provided prior to that evaluation.

TA3.1 and TA3.2: Resources Provided by DARPA

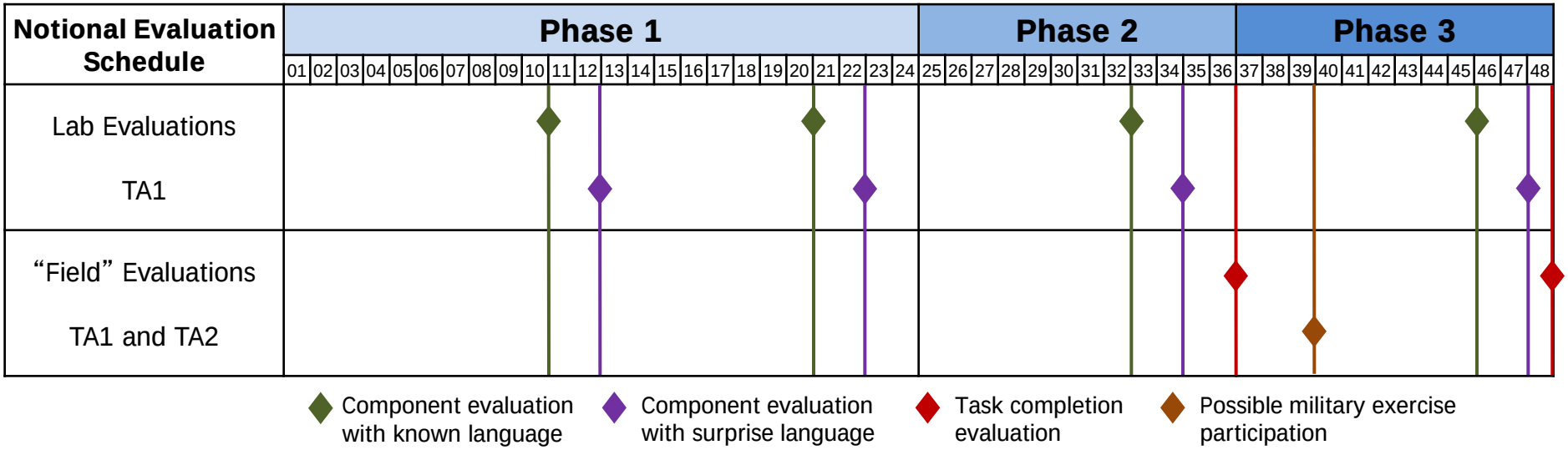
DARPA will provide to all performers:

- REFLEX language packs (beginning of program)
- Applicable BOLT corpora to include:
  - Uzbek data similar to LRLP (release of BAA)
  - Turkish data similar to IL (release of BAA)
  - Hausa data similar to LRLP (beginning of program)
  - Mandarin data similar to IL (beginning of program)
  - Mandarin data from BOLT – as applicable (beginning of program)
  - Arabic data from BOLT – as applicable (beginning of program)

Performance Evaluation: Overview

How will we know if LORELEI has succeeded?

- Lab Evaluations
  - Linguistic component evaluations with known languages and "surprise" languages
  - Conducted by NIST, open to the research community at large
- "Field" Evaluations
  - Task completion evaluations with "surprise" languages and operational scenarios
    - A/B Testing: compare accuracy of results produced using today's workflow to accuracy of results using LORELEI (users will be military analysts and/or planners)
  - Also, trial participation in Humanitarian Assistance and Disaster Relief exercises or other military exercises (e.g. AFRICOM Western Accord)
- **Operational success: LORELEI-equipped non-linguist mission planners are able to produce answers that fall in the human expert performance range.**



Note: Not every evaluation will necessarily be conducted every year

Approved for Public Release, Distribution Unlimited

Performance Evaluation: LORELEI Lab Evaluations

- Require participation by LORELEI TA1 researchers in DARPA-sponsored NIST open community natural language processing evaluations
  - Task: natural language processing component tasks
  - Data: supplied by TA3.2
  - Test conditions:
    - 1: Low-resource language announced in advance, no time constraints (possibly the previous year's surprise language)
    - 2: "Surprise" low-resource language announced at evaluation time. Technologies will be assessed one day, one week, and one month after release of the evaluation language and resources
  - Calibration: compare against manually-produced ground truth
  - Evaluation: example relevant NIST tasks
    - Machine translation — metric: BLEU/METEOR, HTER, HyTER, adequacy
    - Named-entity/place-name extraction — metric: Precision, Recall, F1
    - Entity linking — metric: Precision, Recall, F1
    - Topic/keyword detection and classification — metric: Precision, Recall, F1, term weighted value
    - Event detection — metric: Precision, Recall, F1

Performance Evaluation: LORELEI "Field" Evaluations

- Compare results using LORELEI TA1 + TA2 to current workflow and to ground truth
  - Task: analysts/planners produce situational awareness information
    - Example output: top ten hotspot locations (one-day checkpoint)
    - New "surprise" language used for each evaluation
  - Data: archived incident-related data, live military exercise data
  - Test conditions:
    - A: Group of linguists using today's workflow
    - B: Group of non-linguists using LORELEI
  - Calibration: establish human performance range in today's environment by analyzing responses from five analysts/planners
  - Evaluation: compare group A vs. group B responses
    - Adequacy metric: whether group B responses are within the range of group A responses
    - Accuracy metric: how group A and group B responses compare to ground truth

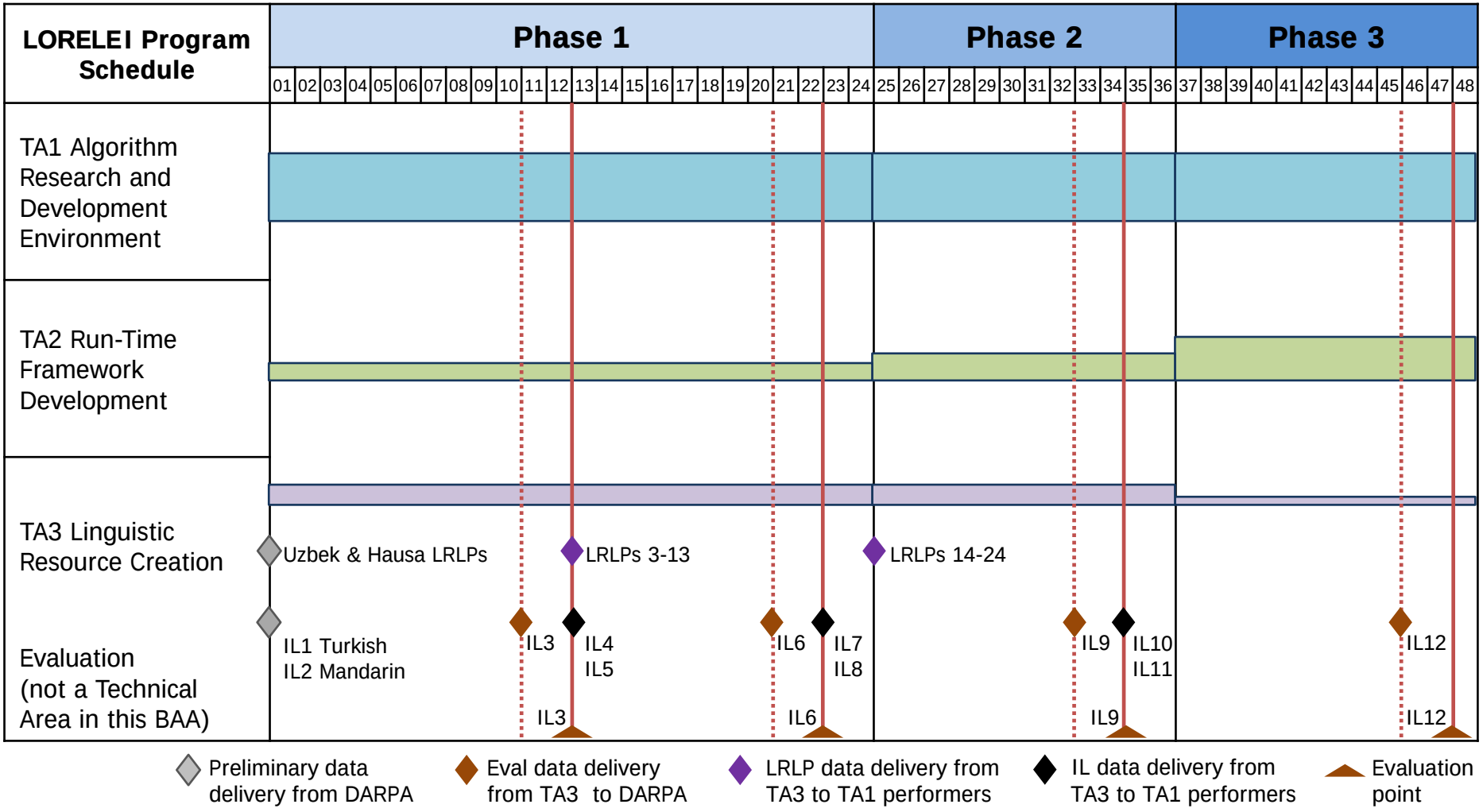
Note on Evaluation: The purpose of the performance evaluations is not to serve down-select decisions, but to inform the adjustment of research directions in the program. Evaluation results will be used as a feedback mechanism to determine which technologies require further improvement and which are mature enough for implementation.

Performance Evaluation: Resources Provided by DARPA

- Scenario model in English
  - Scenario model specific to incident type
- LORELEI representative language packs (LRLPs)
  - Resource sets for typologically representative languages.
- Native-speaker linguistic informant
  - Access to a certain number hours with a native-speaker linguistic informant
- Expert theoretical linguist
  - Access to unlimited hours of one on-site expert theoretical linguist

Example Generic HADR Terms		
Class	**Subclass**	**Examples**
Resource Locations	Airports, Ground Transportation, Seaway, Medical	*Runway, Tarmac, Highway, Passage, Cargo, Clinic*
Location Reference	Community, Geographic, Directional	*School, Market, Ridge, Surround, North*
Resource	Material handling equipment	*Forklift, Crane, Tractor*
Conditions	Transportation, Structure, Weather, Environmental	*Impassable, Suspension, Twisted, Windy*
Infrastructure	Electrical, Water, Communication, Gas	*Connected, Power, Lines, Methane*
Economy	Agriculture, Commerce, Fishing, Manufacturing	*Dust, Crop, Commodity, Assembly, Tide*

Expected Program Schedule



Proposal Information: Scope

- Each proposal must fit exactly one of these scope descriptions:
  - Part or all of TA1.1
  - Part or all of TA1.2
  - Part or all of TA1.3
  - Some combination of the above
  - ALL of TA1 (meaning all of TA1.1, TA1.2, TA1.3, *AND* TA1.4)
  - ALL of TA2
  - ALL of TA3.1
  - ALL of TA3.2
- Selection for multiple technical areas:
  - Proposers selected for TA3.2 can also be selected for TA3.1, but proposers selected for TA3.2 cannot be selected for any portion of TA1 or TA2, as a prime, subcontractor, or in any other capacity from an organizational to individual level.
- During the course of Phase 1 of LORELEI, TA1.1, TA1.2, and TA1.3 performers are expected to demonstrate value to the overall program goals.
  - The preferred method for demonstrating value is to work with a TA1.4 performer to integrate TA1.1, TA1.2, or TA1.3 algorithms, knowledge, resources, or other results into a TA1.4 performer's workflow. If a proposer for TA1.1 – 1.3 is seeking a cooperative agreement and chooses not to provide their algorithms, knowledge, resources, or other results to the 1.4 performers, the proposal must identify how the team will demonstrate value to the program.
  - TA1.4 performers are expected to be receptive to new ideas, algorithms, and/or resources from other TA1 performers over the course of the program.

Proposal Information: Award Structure

DARPA anticipates the following award structure:

- TA1 - multiple small research awards for part or all of TAs 1.1, 1.2, and 1.3. Additionally, DARPA anticipates multiple large awards addressing the entire Technical Area 1 (including TA1.4)

- TA2 - one large award. Work will begin at a low level in Phase 1, consisting mainly of planning and coordination, to include definition of application programming interfaces (APIs) and coordination with TA1 performers as to specific incident language processing tools. In Phase 2, it is anticipated that the level of effort for TA2 will expand several fold to include LRTF prototype building, as well as integration and coordination with transition partners. In Phase 3, it is anticipated that the work of TA2 will increase still more, as prototypes are tested in transition partner spaces.

- TA3 - multiple awards where the brunt of TA3.1 data resources are delivered by the end of year 2 and deliverables drop off significantly in years 3 and 4. There may be multiple partial awards for TA3.1 and/or TA3.2.

Proposal Information: Program Exclusions

- DARPA is ***NOT*** seeking proposals for the following tasks, technologies, or approaches:
  - Performance evaluation
  - Research addressing a single language or even a family of languages
    - Instead, specialists in language families are encouraged to enter teams that provide global coverage in the aggregate
  - Evolutionary improvements to the existing state of practice
  - Incremental improvements to existing Machine Translation systems
  - Any human language technology primarily reliant on crowdsourcing for language knowledge
  - Any human language technology primarily reliant on (English/incident-language) parallel corpora for language knowledge
  - Technologies primarily addressing English or any other specific language

Proposal Information: Schedule

- BAA Posting Date: October 30, 2014 (ET)
- Abstract* Due Date: November 21, 2014, 12:00 noon (ET)
- Proposal Due Date: January 9, 2015, 12:00 noon (ET)
- BAA Closing Date: January 9, 2015, 12:00 noon (ET)
- Projected Kickoff: May 2015

Submissions received after due dates will not be reviewed.

*Proposers are highly encouraged to submit an abstract in advance of a proposal to minimize effort and reduce the potential expense of preparing an out of scope proposal.

Reference Documents - Primary

- The Joint Chiefs of Staff. (2014). *Foreign Humanitarian Assistance*. (Joint Publication 3-29). Washington, DC: Joint Chiefs of Staff.
- U.S. Department of Defense. (2011). *Department of Defense Support to Foreign Disaster Relief (Handbook for JTF Commanders and Below)*. (Document number: GTA 90-01-030). Washington, DC: Government Printing Office.
- U.S. Agency for International Development (USAID) & the USDA Forest Service. (2005). *Field Operations Guide for Disaster Assessment and Response*. Washington, DC: USAID Bureau for Democracy, Conflict, and Humanitarian Assistance.
- U.S. Department of the Army & United States Marine Corps. (2013). *Multi-service Techniques for Civil Affairs Support to Foreign Humanitarian Assistance.* (Army Techniques Publication 3-57.20/Marine Corps Reference Publication 3-33.1C). Washington, DC: Departments of the Army and Marine Corps.

Reference Documents - Secondary

- U.S. Africa Command. (2014). *United Assistance Intelligence Summary 26 Sept. 2014*.
- U.S. Department of the Army. (1998). *Intelligence officer's handbook*. (Army Mission Planning Intelligence Manual Field Manual 34-8-2). Washington, DC: Headquarters, Department of the Army.
- U.S. Congressional Research Service. (2012-01-23). *Building civilian interagency capacity for missions abroad: key proposals and issues for congress*. (CRS Report for Congress R42133. Serafino, N. M., Dale, C., & Towell, P.). Washington, DC: Congressional Research Service.
- Doman, D. M., Balri, J. A., Napierala, M. A., & Cho, M. S. (2011). Do plans and execution agree in a humanitarian medical mission? *Journal of Surgical Orthopaedic Advances*, *20*(1), 67-73.
- U.S. Department of the Army. (1994). *Intelligence preparation of the battlefield*. (Army Field Manual 34-130). Washington, DC: Department of the Army.
- U.S. Department of Defense. (2014, June 23). *Humanitarian and Civic Assistance (HCA) activities*. (DOD Instruction 2205.02). Washington, DC: Under Secretary of Defense for Policy (USD(P)).
- Daly, MAJ Charles. *Humanitarian Assistance and Disaster Relief Communications for the 21st Century*. Thesis. Naval War College, 2007.
- U.S. Department of the Army, United States Marine Corps, & U.S. Department of the Navy. (2014). *Insurgencies and countering insurgencies*. (Army Field Manual 3-24/Marine Corps Warfighting Publication (MCWP) 3-33.5). Washington, DC: United States Marine Corps
- Marine Corps Intelligence Association, Inc. (2014). Intelligence support to HA/DR operations. *INTSUM Magazine*, *25*(4) Spring-Summer 2014 full issue.
- The Joint Chiefs of Staff. (2009). *Joint intelligence preparation of the operational environment.* (Joint Publication 2-01.3). Washington, DC: Joint Chiefs of Staff.
- The Joint Chiefs of Staff. (2012). *Peace operations*. (Joint Publication 3-07.3). Washington, DC: Joint Chiefs of Staff.
- The Joint Chiefs of Staff. (2013). *Operations in chemical, biological, radiological, and nuclear environments*. (Joint Publication 3-11). Washington, DC: Joint Chiefs of Staff.
- The Joint Chiefs of Staff. (2011). *Military information support operations*. (Joint Publication 3-13.2, 7 January 2010, Incorporating Change 1, 20 December 2011). Washington, DC: Joint Chiefs of Staff.
- The Joint Chiefs of Staff. (1995). *Joint doctrine for military operations other than war*. (Joint Publication 3-07). Washington, DC: Joint Chiefs of Staff.
- The Joint Chiefs of Staff. (2013). *Multinational operations*. (Joint Publication 3-16). Washington, DC: Joint Chiefs of Staff.
- Lidy, A. M., & Kunder, J. (2005). Large-scale military humanitarian assistance. In *Handbook on the analysis of smaller scale contingency operations in long term defence planning* (NATO RTO-TR-SAS-027, Symposium Paper 7).
- Mogilevsky, LCDR Phillip. *Optimizing Transportation of Disaster Relief Material to Support U.S. Pacific Command Foreign Humanitarian Assistance Operations*. Master's thesis. Naval Postgraduate School, 2013.
- Kauffeld, B.D. for the Peacekeeping and Stability Operations Institute. (2014). *USAID & DOD: Analysis and recommendations to enhance development-military cooperation*. Carlisle Barracks, PA: United States Army War College Press.
- Pavich, MAJ Todd. *Using the Air Force to Conduct Humanitarian Assistance in a Hostile Environment*. Master's thesis. U.S. Army Command and General Staff College, 2004.
- Yale-Tulane ESF-8 Planning and Response Network. (2014). *Special report: West Africa – Ebola 2014 - 19 September 2014 Update*. New Haven, CT: Yale-Tulane ESF-8 Medical and Public Health Team.

www.darpa.mil